Sadaf Munshi-DEL-2009

ARCHIVE OF ANNOTATED BURUSHASKI TEXTS (Ethnologue Code: BSK)

Project Description

1. Project Overview

The primary goal of the project is to create a linguistically analysed and searchable Archive of Annotated Burushaski Texts (AABUT). The database for AABUT will consist of 15-20 hours of high quality digitized audio and video recordings of oral literature from different, significantly threatened, regional varieties of Burushaski, viz., Hunza, Nagar, and Yasin in Pakistan and Srinagar in India. Following the latest recommendations for language documentation (Bird & Simons 2003; Simons & Bird 2006; E-MELD 2005a-e and 2006a-b; OREL 2009), the project team will collect, analyze, annotate, and archive Burushaski oral literature in various genres including: traditional stories and mythical legends, historical accounts, personal narratives, formal speeches, conversations, poems, speaker-to-speaker discussions of data, and metadata. About 8-10 hours of the audio and/or video materials will be provided with time-aligned transcriptions, multi-tier annotations, morpheme-to-morpheme and free translations into English. A selection of smaller edited texts will be transliterated into Burushaski with Urdu translations. A second outcome of the project will be pedagogical materials for promoting language revitalization; this will be done considering the community's strong interest in this aspect and the reviewers' recommendations for revising the earlier unfunded version of this proposal. The pedagogical materials will consist of: introduction to the Burushaski alphabet, wordlists, parts of speech, basic grammar, and smaller texts (e.g. short stories, poems, songs, etc). Transliteration in Burushaski will be done using a modified version of the Perso-Arabic (Nastāliq) script based primarily on the Urdu writing system. The documentation materials, including data, description of data, texts and translations, sound files and video recordings, and metadata, will be archived at the University of North Texas (UNT) Digital Collections Library. Access to the materials will be restricted as determined and requested by community members, and materials will be disseminated to the community in print, online and as other media, such as, CDs, DVDs, etc.

2.0. Language Context: Need for Documentation

2.1. Sociolinguistic and demographic background: Based on the government census figures of 1981, Backstrom (1992: 36) estimates the total number of Burushos in Pakistan as 55,000-60,000. Of these, 18-22,000 live in Hunza, probably slightly more than that in Nagar, and about 15,000 or more in Yasin. Hunza, Nagar, and Yasin valleys are situated in the Gilgit District of Northern Areas in Pakistan. *Ethnologue* (2005) reports 87,049 speakers of Burushaski in

Pakistan. There are significant dialectal differences between the Yasin variety (Werchikwar), on the one hand, and Hunza and Nagar on the other (Varma 1941). About 300 speakers of Burushaski live in Srinagar, the capital of Jammu & Kashmir, India, who refer to their language as *miśa:ski* 'our language' or *Khajuna*. This variety, the JKB, separated from the Nagar variety in 1890-1 (Hassnain, 1978), and is a distinct variety of Burushaski exhibiting systematic differences with other varieties (Munshi 2006, 2009).

Burushaski is spoken in a region home to speakers of several language families: Indo-Iranian, Tibeto-Burman, and Altaic (Anderson 1997). It has been greatly influenced by languages such as Persian, Arabic, Urdu, Khowar, Shina, Wakhi, Balti and Kashimiri. A majority of the Burushos (speakers of the Burushaski language) are multilingual in their native language and at least one of the other regional languages, e.g., the Indo-Aryan Urdu, Shina, Kashmiri, and Khowar, and the Tibeto-Burman Balti (Anderson 1997, Kohistani & Schmidt 2006, Munshi 2006). Among these, Urdu has a special status in that it is the *lingua franca* of the region and the language of literacy (Alam 2003). Just like speakers of other minority languages, dominance of Urdu has resulted in a strong push for a majority of Burushos to shift to Urdu (Saxena & Borin 2006). With greater means of mobility, more and more people have chosen to move to bigger cities for education and employment. As a result, they shift to using Urdu as their primary language. Thus, imperfect knowledge of the language is very common and fluency in Burushaski among the second and third generation is on a rapid decline. Equally strong factors resulting in language shift are lack of institutional support and cultural homogenization through education and media.

2.2. Genetic classification – language isolate: Many studies on Burushaski deal with attempts to trace its linguistic origins (cf. Toporov 1970, 1971; Bengston 1991, 1992, 1996 [1998]; Tuite 1998; Čašule 1998, 2009; and others). For example, according to Bengtson, Burushaski would belong to a "Macro-Caucasic" family under the "Dené-Caucasian" macrophylum (Bengston 1991, 1992, 1996 [1998]) – a proposed, transcontinental branch consisting of Basque, languages spoken in Daghestan, North-West Caucasian languages, Na-Dene, and Burushaski. Čašule (1998, 2009) has attempted to derive links between Indo-European, more specifically its Paleo-Balkanic branch, and Burushaski. So far, none of these studies provide conclusive evidence for a genetic relationship between Burushaski and an existing language. Therefore, the language is still considered an *isolate***.**

2.3. Typological features: Some typological features of Burushaski which make it a very interesting language are:

a. Presence of a rich inventory of *retroflex* consonants (consonants pronounced with the tip of the tongue curled up) such as *ċ*, *ċ*^{h}, *ṣ*, *ż*, and *y*, and uvular stop and fricative consonants such as, *q* and *γ*. Some, but not all, of these sounds are found in its neighboring languages, such as Dardic Khowar, Shina, Kohistani; these consonants are absent in Urdu/Hindi, Kashmiri, Punjabi, and many Indo-Aryan/Indic languages spoken further south and east (Note: Retroflex consonants are claimed to be due to the influence of a

Dravidian substratum in the area. Uvular stop and fricative consonants in some IA languages, such as Urdu, are only present in lexical borrowings from Persian/Arabic).

b. A huge number of different plural marking suffixes. For example, *-muc, -(v)anc,-einc, -(im)iŋ, -išo, -(ič)aŋ, -v, -(y)o,* and *-koyo,* to name a few.

c. A very rich and complex agreement system; Ergative, Absolutive, as well as Dative agreement is possible; agreement features for Ergative/Dative as well as Absolutive arguments simultaneously expressed on the verb when both are present in a sentence.

d. Distinction into inherently possessed nouns (such as words for body parts and kinship terms) as opposed to non-inherently possessed nouns (including most other nouns) – a feature absent in most (if not all) neighboring languages.

e. A four-fold classification of nouns: [+human] vs. [-human]; among humans, [male] vs. [female]; and among non-humans, [+concrete] vs. [-concrete]. These are differently expressed in nouns and verbs for agreement patterns – another feature not found in its neighbors.

f. Presence of an overt Indefinite Article. The only other language in the same linguistic area which has this feature is Kashmiri. Consider, for instance, Burushaski *dasin-an* and *huk-an* vs. Kashmiri *ku:r-ah* and *hu:n-ah* meaning 'girl-Indefinite; a/any/some girl' and 'dog-Indefinite; a/any/som dog' (respectively).

g. Presence of certain postpositions which do not necessarily have exact parallels in its neighboring languages. For example, postpositions for the adverbials of Time/Duration and Space/Location are different.

h. A unique way of signifying kinship terms, very different from the neighboring languages and cultures, determined by the gender of both the qualifier/possessor and the qualified/possessee. For example, *eċu* '(his) brother', *moċu* '(her) sister', *mu:lus* '(her) brother', and *yas* '(his) sister'.

2.4. Available materials: A substantial amount of published literature dealing with Burushaski language and society in Pakistan exists (See Tiffou 2000). Many linguistic studies are dated (e.g., Biddulph 1880, Leitner 1889, Zarubin 1927, Lorimer 1935-8, Varma 1941) and contain various inaccuracies in the linguistic analyses. Some of the recent studies on the language are: Anderson (1997, 2003, 2007), Anderson & Eggert (2001), Berger (1974, 1998), Tiffou & Pesot (1989), Tiffou (1993, 1999), Skyhawk (2003), and Willson (1999). The primary documenation references on the language are Berger (1998) and Tiffou (1999) on the Hunza-Nagar (predominantly Hunza) and Yasin dialects of Burushaski (respectively). The studies, however, are in German and French, and, therefore, inaccessible to Burushaski speakers who are literate in Urdu and English. The text collections are not linguistically analyzed and only broad translations are provided.

Recently, scholars working with a local non-government body, the *Burushaski Research Academy* (BRA), have produced some pedagogical materials and a Burushaski-Urdu dictionary based on the Hunza dialect (Nasir Hunzai n.d.-2003). Because of limited resources and lack of appropriate training, such efforts are yet to produce substantial outcomes. Many of these materials have substantial room for improvement by trained linguists and language experts.

Secondly, because the organization primarily focuses on serving and promoting the Ismaili community in Hunza, other Burushos are reluctant to participate in their activities.

2.5. Need for documentation: As a genetically isolated and a typologically very interesting language, Burushaski makes a strong case for documentation and preservation. There are extensive dialectal differences in all areas of grammar which have not been addressed in the previous studies. In order to fully understand some of the most complicated areas of Burushaski grammar and typology, such as, grammatical case and agreement, studying these differences is very important. A linguistically analyzed, cross-dialectal study with a large corpus of first-hand data will bring new insights into some of the theoretical aspects of the language. A comprehensive cross-dialectal study of the language is the need of the time for achieving a more rigorous historical and comparative reconstruction of the language and for finding the right place for Burushaski in the grand tree of languages – a topic of great intellectual debate in many previous studies.

Burushaski is primarily preserved orally and literacy in the first language is practically non-existent. The survival of the different varieties of Burushaski is greatly threatened by increasing and rapid language shift. The langauge has a very rich story-telling tradition which is yet to be fully explored. Many Burushos have expressed a strong need for the documentation and preservation of Burushaski oral literature which is feared to be lost when the older generation passes away. Despite a number of previous documentation efforts, the speakers have been unable to avail themselves of the materials because most of these are in foreign languages which the Burushos are not literate in. A substantial amount of available materials on Burushaski are based on the Hunza dialect, and the other dialects, viz., Yasin, Nagar and Srinagar, are especially in need of documentation.

A major problem faced by scholars working on Burushaski is the absence of a widely-accepted writing system for the langauge. Local rivalries have resulted in a number of competing writing systems and an official consensus about a standardized script is yet to be made. For example, only in Hunza, at least two variant writing systems exist which were proposed by Nasir Hunzai (of BRA) and Ghulam Hunzai (See Ghulam Hunzai 1990) – both highly respected members of the community who have composed a number of poems and other literature in the language. Furthermore, there are a number of problems with the proposed writing systems which need the attention of a trained linguist who is also familiar with the writing systems of the region.

Considerring these factors, there is a call for a new and broad-based documentation project which will address the shortcomings of the previous documentations as well as provide fair representation to all the dialects. This can only be acheived by a comprehensive cross-dialectal study involving community members and language experts. As we will see in the following sections, all the infra-structure for such a project is already in place, except for funding.

3.0. Project Participants

This project proposes to put together a research team which will be composed of: the principal investigator (PI), two research assistants, and several native speaker consultants.

3.1. Principal Investigator: Dr. Munshi will conduct the overall management of the project and coordinate and conduct fieldwork in Yasin, Nagar, Hunza and Srinagar. She is uniquely qualified to carry out this project for several reasons:

The region where Burushaski is spoken is politically sensitive and also traditionally conservative. It is, therefore, very important that the researcher knows the region well. As a native of Srinagar, Dr. Munshi is familiar with the cultural traditions of all Burushaski-speaking areas. She has sought assistance from and established contacts with several highly motivated and sophisticated Burushaski speakers as well as community organizations and educational institutions across different dialects of Burushaski. Dr. Munshi has a thorough understanding of the Burushaski grammar and a working proficiency in the Srinagar variety. She shares at least two languages with the community, viz., Kashmiri (Srinagar) and Urdu. She has a strong background in other important languages, including, Sanskrit and Persian. She also knows the writing systems used in the region, which include different modifications of the Perso-Arabic (Nastālīq) and Devanagri scripts. An additional advantage is her gender which gives her easy access to female speakers in a conservative Muslim community which does not welcome male researchers inside the household unless they are closely related to the family. As the speech of many of Burusho females is less influenced by contact than their male counterparts, access to data from female speakers is a must for a documentation project.

This project is the second step in Dr. Munshi's continuing work on the documentation of Burushaski. The first step in this regard was her doctoral dissertation, submitted to the University of Texas at Austin (UT Austin) in 2006 (supervised by Dr. Anthony C. Woodbury). The study provides a structural description of JKB – a previously undocumented variety of Burushaski spoken in Srinagar, and analyses the various forms of linguistic interference and language change in this dialect (Section 6). JKB is a direct descendant of the Nagar dialect of Burushaski in Pakistan – a dialect which has been fairly underrepresented in most previous studies. Another outcome of Dr. Munshi's work on Burushaski is a paper entitled "Contact-induced language change in a trilingual context: the case of Burushaski in Srinagar" to appear in *Diachronica* (2010). The study makes an important contricution to studies on language contact. Data from this study indicate that lexical borrowing and structural borrowing are two different types of contact phenomena which can occur independently of each other; the two processes are influenced by different sociolinguistic factors which may interact in different ways in different contact situations resulting in different outcomes. Munshi has also presented a number of papers on different aspects of Burushaski language and documentation at various academic conferences.

Over the past six years, the PI has conducted intensive linguistic fieldwork (12-13 months) with native speakers of Burushaski in India (Srinagar) and in the United States (Austin, Houston, and New Jersey). She has created a corpus of Burushaski spoken in Srinagar which includes: audio and video recordings of 12 stories (children's stories, folk tales, myths, and religious stories), 8 personal narratives, 18 natural conversations, 1 oratory speech, a number of poems, and about 200 pages of hand-written fieldnotes (excluding the transcriptions of the texts listed). The data were collected during five fieldtrips funded by: a UT Austin Liberal Arts Graduate Research Fellowship ($3000, 2003), an NSF Doctoral Dissertation Improvement Grant ($12,000, 2004), and three UNT Faculty Research Grants ($5,000 each in 2007, 2008, and 2009). A significant amount of the data have been digitized, transcribed, translated and linguistically analyzed.

The proposed project fits in with a strong and established research program at UNT in the PI's department of Linguistics and Technical Communication (LTC) of work on endangered languages as evidenced by the number of faculty members at LTC who have received federal funding in recent years to document and preserve endangered languages spoken in different areas of the world. These include: Dr. Shobhana Chelliah in 2008 on Lamkang –a Tibeto-Burman language spoken in Manipur in India, Dr. Timothy Montler in 2007 on Kllalam – a Native American language of the Salishan language family/group spoken in northwest Washington, and Dr. Willem de Reuse in 2006 on Western Apache spoken in central Arizona. The project also complements an upcoming event at the University of North Texas organized by the PI (in collaboration with Dr. Chobhana Chelliah) – the 28th meeting of the *South Asian Languages Analysis* roundtable (SALA XXVIII) to be held at UNT in October 2009 (www.sala.unt.edu). One of the focal areas of the conference is endangered languages of South Asia – a topic which has received relatively little attention at previous academic meetings (Note: South Asia is a repository of many languages which are currently under great threat from the influence of more dominant languages. It is, therefore, imperative that large-scale documentation studies are conducted and the languages are preserved before a point of no return when there are too few native speakers left to work with).

3.2. Other Research Staff: A trained native speaker ("research assistant") and a graduate student ("technical assistant") will assist the PI on the project.

Research Assistant: Piar Karim – a lecturer of English at the Agha Khan University (AKU, Karachi, Pakistan) and a native Hunza Burusho, with a Master's degree in English Linguistics and Literature, will be the primary research assistant of the PI. Karim is a fluent speaker of Burushaski with advanced proficiency in Urdu and English (written and spoken) and is well-versed in the Perso-Arabic script. Karim is extremely highly motivated to work on the project; he is a very useful and strong liason between the PI and various local Burushaski scholars and native speakers in Hunza, Nagar, and Yasin with whom he has discussed the project at length.

Karim has indentified a number of potential native speaker consultants in Hunza, Nagar and Yasin and collected digital audio recordings of several speech samples from them during summer 2009. He has a fair understanding of and experience in aspects of linguistic fieldwork and data collection: use of equipment, data transfer and storage, transcription, morphemic analysis; collecting useful information, such as, demography and metadata; and issues regarding fieldwork ethics, informed consent, and payment of native speaker consultants.

Technical Assistant: A UNT graduate student with a substantial background in programming and web designing as well as experience and training in using FLEx and ELAN will be hired as a technical assistant for the project. The technical assistant will help the team in preparing transcriptions and time-aligned annotations, preparing archive-ready materials, and in setting up the database and managing the archive. A graduate student with such qualifications has already been identified by the PI for the purpose.

3.3. Community Participation and Native Speaker Consultants: The following native speakers have already committed to participation in both aspects of the project – creation of the digital archive (documentation) as well as preparation of pedagogical materials (revitalization):

Yasin: Mohammad Wazir Shafi – a lawyer from Yasin who worked with Tiffou (1999) on Yasin Burushaski and has also written a book *Burushaski Razon* (n.d.) which provides a grammatical sketch of Yasin Burushaski, well known in Yasin because of his contributions to the language; and Sardar Khan – a young Yasin Burusho scholar based in Karachi.

Hunza: Dr. Nasiruddin Nasir Hunzai – chief patron of BRA, who collaborated with Tiffou (1993) and Berger (1998); Shahnaz Salim Hunzai – director of the on-going Burushaski-Urdu dictionary project at BRA; Dr. Mola Dad Shafa – Head Professional Development Center, North, AKU Institute of Educational Development; and Ghulamuddin Ghulam Hunzai – a noted Burusho scholar who has composed substantial literature in Burushaski using a writing system that he proposed.

Nagar: Syed Mohammad Yahyah Shah, an academician known for his social work and scholarly activities; Ahmad Hussain Nagari, a young Burusho with a degree in linguistics; and Ismail Tehseen, a historian with interest in Burushaski and history of Nagar, who is associated with the Nagar Academy Chalt (NAC) – a local non-government body working for the promotion of the language in Nagar.

Srinagar: Raja Safdar Ali Khan (a retired school teacher), Raja Jamsheed Ali Khan and Raja Mehboob Ali Khan (retired government employees) – all three are great story-tellers and have immensely contributed to data collection and analyses over the past six years; Hasina Bano – a

young Burusho woman, with prior experience and training in data collection and analyses; and Saba – a Burusho woman trained in aspects of pedagogy and language teaching.

4.0. Research Objectives and Plan

4.1. Primary Objectives: During the proposed project the PI will coordinate and conduct fieldwork in Yasin, Hunza, and Nagar in Pakistan, and Srinagar in India -- regions where different Burushaski dialects are spoken. The work will be carried out by a team consisting of: the PI, trained native speaker consultants (including a primary research assitant), and a UNT graduate student assistant. The proposed work includes:

- Fieldwork: collecting audio/video recordings of natural linguistic data,
- Training native speakers of Burushaski in methods of documentation,
- Data transfer and storage: transfer of audio and video materials from one form of equipment to another and copying for storage,
- Data transcription using a Unicode compliant font,
- Data analyses: parsing and interlinearization of texts with detailed grammatical information and translation in English,
- Transliteration of texts into Burushaski and their translation into Urdu,
- Development of pedagogical materials, and
- Archiving of materials for documentation.

For documentation and archiving, the team will collect audio and video recordings of 15-20 hours of Burushaski speech samples in different genres including:

- Traditional stories,
- Myths and legends,
- Historical accounts,
- Personal narratives,
- Formal speech,
- Natural conversations,
- Songs and poems, and
- Speaker-to-speaker discussion of linguistic data.

For language revitalization purposes, the team will create the following pedagogical materials:

- An introduction to the Burushaski alphabet along with words and spelling conventions/rules,

- Wordlists: numerals, body parts, kinship terms, color terms, days of the week, weather terms, greetings and expressions,
- Parts of speech: demonstrative pronouns, possession, adjectives, postpositions, conjunctions and complementizers, Number and Gender,
- Basic grammar: noun and verb inflections, and
- A selection of small edited texts, short stories and poems.

Dialectal differences will be provided wherever possible or relevant.

4.2. Documentation Methods: The project team consisting of the PI and trained native speaker consultants will collect 15-20 hours of speech samples from speakers of different Burushaski dialects. Data will be collected from both male and female speakers. While stories, legends, historical accounts, personal narratives, poems and songs will mainly be collected from older speakers, with special care taken to include people whose speech is less affected by contact, natural conversations will involve speakers of different age groups with varying degrees of proficiencies. This will generate data of different types which can be used in many different ways. To the extent possible, the data from the different varieties of Burushaski will be collected in more or less equal proportions. In an attempt to avoid *observer's paradox*, audio/video recordings will mostly be made by trained native speakers in absence of the PI.

For transcribing, analysing and archiving the texts, the project team will follow the latest technological recommendations for language documentation (Bird & Simons, 2003; E-MELD 2005a-e and 2006a-b; and OREL 2009). Audio files will be captured in digital wave format using a solid-state audio recorder. Video-recordings will be made using a high resolution digital video camera. Texts will be linguistically analyzed and annotated with time-synchronized transcription and translation using ELAN – a professional tool for the creation of complex annotations on video and audio resources. For maintaining morphosyntactic analyses and lexical files, we will use FLEx – Fieldworks Language Explorer. The PI has prior experience in using both ELAN and FLEx. Data will be transcribed phonemically using a Unicode compliant font with morpheme-to-morpheme analysis and free sentence translations into English (Appendix 1). Wherever feasible, ethno-linguistic or other culturally or historically relevant information will be included (Appendix 2).

Keeping in view the level of literacy and second language proficiency among Burushaski speakers, a version of the texts will be independently transliterated into Burushaski script and translated into Urdu (Appendix 3). For transliteration into Burushaski, we will use a modification of the writing systems used by the Burushaski Research Academy and by Ghulamuddin Ghulam Hunzai (Appendix 4). The proposed modified version will entail more systematicity and economy by use of fewer new symbols which were added to the base Urdu alphabet – a widely-used modification of the Perso-Arabic script. To the extent possible, the script will avoid conflict with any pre-existing symbols proposed for the language; it will eliminate some of the variant symbols used by different local Pakistani Burusho scholars; and it will also do away with some problematic

symbols in the recently devised Nafees Pakistani Web Naskh Open Type Font (NPWNOT) - BTK – a character-based font claimed to be developed on Unicode standards for Burushaski, Torwali, and Khowar by the Center for Research in Urdu Language Processing (CRULP), Pakistan. (Note: A number of characters in NPWNOT-BTK do not properly represent linkage between adjacent letters in word-medial and final positions; See Hussain & Karamat (n.d.). Therefore, another version of the font called "NPWNOT – modified" was created by CRULP. This, however, caused a new problem because symbols for certain Urdu sounds which are also present in Burushaski have been incorrectly mapped to different symbols for certain Burushaski-specific sounds, thus, leading to the inadequacy of symbols in the font. The problem has been conveyed by the PI to the creators of the font; a possible resolution is still awaited). The proposed changes will be made in consultation with representatives from different dialects.

The archive-ready materials will be produced in the following formats: audio in PCM wav, video in MPEG-2, texts in HTML, Word and PDF, and hand-written documents and (fieldwork) pictures in JPEG. Data will be transferred to a portable external hard drive and a laptop hard drive. For archiving and presentation of the data the PI intends to follow the format used by AILLA – a digital archive of recordings and texts in and about the indigenous languages of Latin America at the UT Austin, which she is also (loosely) affiliated with as a graduate alumni. She is also in consultation with Dr. Anthony Woodbury and Heidi Johnson of UT Austin and partners at AILLA as well as her colleagues including, Dr. Shobhana Chelliah and Dr. Timothy Montler.

All documentation materials, including data, description of data, texts and translations, sound files and recordings, and metadata, will be archived at the Digital Library Collections of the University of North Texas. Preliminary preparations in this concern have already started in collaboration with Cathy Hartman, the Assistant Dean of the Digital and Information Technologies at UNT. Supplemental copies will be made on CDs, DVDs, and in print, and maintained safely by the PI in locked cabinets in her office at UNT. The materials will be disseminated to the Burushaski communities online, in print and as other media (CDs, DVDs, etc.).

4.3. Project work plan:

Phase I (ongoing – August 2010): A substantial amount of data from Srinagar has been collected by th PI. These include: audio and video recordings of 12 stories, 8 personal narratives, 18 natural conversations, 1 oratory speech, 3 poems, and about 200 pages of hand-written fieldnotes for data elicitations. Of these, the PI has completed the transcription of 6 stories, 5 personal narratives, 2 historical accounts, and 2 poems on paper with the help of native speaker consultants in Srinagar and in the United States. Also completed are the transcription, morphemic analysis and free translation of 2 stories, 1 natural conversation and 1 poem on computer. Approximately 7 hours of audio and 3 hours of video recordings of natural speech have been digitized and converted into WAV and MPEG formats. In summer 2009, data

collection in Pakistan began with the help of Piar Karim who visited Hunza, Yasin, and Nagar valleys in Gilgit. So far, Karim has collected digital audio recordings of six stories and mythical legends, a personal narrative, and a number of poems from speakers of different dialects. The work was supported by a UNT intramural Research Initiation Grant (2008-9; $5000).

In May/June 2010, the PI will conduct a 5 week fieldtrip for more data collection in Hunza, Nagar and Yasin in Gilgit. Wazir Shafi (Yasin), Ismail Tehseen (Nagar), and Piar Karim (Hunza) will escort her to native speakers they have previously identified as willing to participate in the project. In July 2010, Dr. Munshi will conduct a 3-week fieldtrip to Srinagar in order to fill gaps in the previous data collections and transcriptions, while Karim will continue data collection in Gilgit until August. This will give an overall fieldwork time worth 15 weeks (4 weeks each in Nagar, Hunza, and Yasin and 3 weeks in Srinagar). Care will be taken to collect fairly comparable amount of data from each dialect. Based on the specializations of different native speakers, the team will collect about 10-20 hours of new speech samples in different genres (See section 4.1 for details). Besides data collection, meetings will be held with language experts and community representatives from various dialects for the finalization of a widely-accepted Burushaski writing system, the groundwork for which has already begun.

Phase II (September 2010-August 2011): In September 2010-January 2011, Karim will join the PI at UNT for data transcription, translation, and analyses. A large amount of data and fieldnotes will be entered into FLEx to add to the corpus. The technical assistant will undertake time-aligned annotation of the audio and video materials and prepare language files for morphological and syntactic analyses which will be done by the PI and Karim. The period of February-May 2011 will mainly be utilized for the morphological and syntactic analyses of the texts by the PI and the transliteration of selected texts into Burushaski and their free translation into Urdu by Karim – these will be later checked by both together. The technical assistant will conduct the next phase of time-aligned annotation, and start preparation of archive-ready materials, cataloging and management of data and metadata. During May-August 2011, the PI will conduct a 2 month fieldtrip with Karim to discuss materials, transcriptions and translations with native speakers of different dialects, and fill gaps in previous data collections. Pedagogical materials will be developed throughout the year in consultation with specialists at UNT and with the help of trained native speakers and Burushaski scholars including: Shahnaz Hunzai (Burushaski Research Academy), Ghulam Hunzai (a language expert from Hunza), Ismail Tehseen (Nagar Academy Chalt), Wazir Shafi (a language expert from Yasin), Mola Dad Shafa (AKU, Institute for Educational Services), and Safdar Ali (a language teacher and native Srinagar Burusho).

Phase III (September 2011 - August 2012): During September-December 2011, Munshi and Karim will continue the transcription, translation and analyses of the data at UNT. The technical assistant will undertake the next phase of time-aligned annotation of texts, and preparation of archive-ready materials. During January-April 2012, Karim will conduct the next phase of transliteration of the data into Burushaski and their translation into Urdu and the PI will continue the morpho-syntactic analyses, while the technical assistant will work on the entering of data transcribed by Karim, preparing archive-ready materials, cataloguing of data and management of metadata. Webdesigning will be done by the PI in conjunction with the Technical Writing faculty at UNT. Development of pedagogical materials will be undertaken thoughout the year with the help of native speakers. During May-June 2012, the PI will make a final 3-4 week

fieldtrip to clarify any doubts in previous transcriptions, analyses, translations and transliterations. The rest of the summer (until August 2012) will be devoted to finalizing the project, completing the metadata, and depositing the materials to the UNT Digital Collections Library for archiving.

Note: Despite the fact that Northern Areas is a relatively peaceful region as compared to other areas of Pakistan, there is a possibility that the political situation is not conducive for the PI to travel to the country. Under such circumstances, she will invite Karim to UNT in summer 2010 for one month and provide him the required training, equipment and guidance to conduct fieldwork in Pakistan with the assistance of Shahnaz Hunzai and Ismail Tehseen.

4.4. Ethical aspects of the project: Ethics in fieldwork is a topic of interest in many studies on linguistic fieldwork (Himmelmann 1998, Rice 2004, Childs 2007, Bowern 2008, Chelliah and de Reuse (in preparation)). Some issues that a language documenter has to worry about before collecting data relate to what we call the "informed consent" and decisions related to the inclusion or exclusion of the data and their publication. Several questions emerge in the context of working with a community such as that of the Burushos (Munshi 2009). For example: 1) whether the format of the informed consent should be oral or written, 2) should the researcher have any obligations to the community after having sought informed consent, 3) who will be the right person to seek "consent" from, and 4) does seeking informed consent give the researcher the right to access the data sought for?

Burushaski is spoken in a politically sensitive region and people are often hesitant to provide any kind of written commitment. Thus, seeking oral consent is the only alternative. However, there is a problem with this: oral consent is often not taken very seriously, people may not actually consent to what you think they consented for, or they may disagree later. Here is a community where women hardly make their own decisions; it is not, therefore, sufficient to have sought oral consent from them even if they are adults. As a traditional and conservative Muslim community, publishing data which expose women to strangers in a way not acceptable to the elders, may upset the community practices and put some individuals in a socially vulnerable situation. These are some of the very sensitive issues related to fieldwork ethics in the region that the PI is fully aware of. Keeping these issues in mind, the PI will use both the conventional method of seeking informed consent as well as her personal knowlegde about the community's cultural practices in making the materials available for use. Thus, depending on various socio-cultural sensitivities, certain texts will be marked with various degrees of restrictions. Other conventional "ethics" practices will also be followed. Thus, names will be used to acknowledge authorship of materials. No consultant's name will be suppressed, because doing so would be denying the author's right on the material, weaken the data source, and not acknowledge people's participation in the project. Confidentiality of the data and personal information will be made explicit and published only in a form acceptable to the consultants. Pseudonyms will be used when speakers do not want their names published. Any personally identifiable information which the participants want to be kept confidential will be coded and kept in a locker in strict

possession of the PI. Paper records with identifiable information will be stored in locked cabinets and security codes will be assigned to any computerized records.

Note: Regarding the ethics issues, an application for initial review was submitted to and approved by the Institutional Review Board of the University of North Texas. In accordance with 45 CFR Part 46 Section 46.101, the study was determined to qualify for an exemption from further review (Application # 09149; date of approval: April 6, 2009).

5.0. Project Significance

Intellectual Merits: Burushaski is a language isolate, and, thus, a language representing millennia of independent development; attempts to preserve this relic language of the past are an important contribution to different fields in social sciences. A wide range of text types resulting from the project will enable researchers to analyze discourse in its many different aspects and help contextualize verbal art within everyday linguistic resources. By uncovering a rich Burushaski story-telling culture, the project will make an important contribution to the fields of linguistic and cultural anthropology. The project outcomes will be of interest to historians and social scientists by providing materials consisting of orally-preserved history of the Burushos. The data will be useful to sociolinguists interested in the study of language use and language variation. The database will provide useful information about language contact and contact influence in different dialects of Burushaski, and, therefore, be of interest to contact linguists. As a first study of its kind on Burushaski dialectology, the data will provide deeper understanding about the historical development of the language and can be used by historical linguists for a more rigorous internal reconstruction as well as for comparative studies based on the dialects; it will, therefore, be helpful in shedding light on the position of Burushaski in the grand tree of languages. The data will help unravel some of the most complicated areas of Burushaski grammar and typology begging further investigation. No previous study on Burushaski provides a linguistic analysis of texts in the proposed manner; by providing a corpus of high quality original documentation materials in different media – audio, video and written, the project will be an important contribution to the ongoing developments in documentation research. The project includes a training component involving native speakers, thus, resulting in a community-based approach to documentation and revitalization – highly recommended in the recent times. The lexical entries, texts and translations will eventually feed into a multi-lingual cross-dialectal dictionary of Burushaski – a future goal.

Broader Impacts: The proposed project will document different, significantly endangered, varieties of Burushaski. It will contribute to the documentation and revitalization of Burushaski before reaching a point of extreme endangerment. The materials will be accessible by a large community and people of different interests including linguists, historians, cultural anthropologists, as well as native speakers, language teachers and educators. Unlike previous documentations which were largely in German and French, and, therefore, inaccessible to Burushos, in this study the

translations will be done in Urdu and English – languages of wider communication and languages of literacy in the Burushaski-speaking regions. The project will help establish a widely accepted orthography for Burushaski – an issue that has created problems for many previous scholars working on the language. The materials will promote native language literacy and enable the Burushos to read their own literature with the help of languages in which they are literate. By receiving a considerable representation, speakers of all dialects, especially Nagar, Yasin and Srinagar which have been fairly under-represented in previous literature, will be invested with academic respectability, thus filling a gap in the current documentation of Burushaski. By training and involving community members, the project will address various community concerns regarding the language and its preservation and/or revitalization.

6.0 Results from Previous NSF Support

A recent study by the PI, Munshi (2006), has received NSF funding the details of which are:

(a) **NSF award** -- 0418333; Period: 2004-2006; Amount: $12,000

(b) **Title**: "Language Contact and Change in J & K Burushaski" (Note: The title was later changed to "Jammu and Kashmir Burushaski: Language, Language Contact and Change").

(c) **Summary of Results:** The award was used to conduct an eight-month fieldtrip to Srinagar to collect data for the dissertation. The study provides a structural description of JKB – a previously undocumented variety of Burushaski spoken in Srinagar, and analyzes the various forms of linguistic interference. It covers the various linguistic consequences of contact such as: borrowing, innovation, and simplification of linguistic features characterizing JKB. Changes are studied at lexical, phonological, and morpho-syntactic level. Dr. Munshi's synchronic description of the grammar is concerned with the structural properties of the language. Grammatical description is preceded by an introduction of various speech forms in context which emphasizes the importance of a discourse-centered approach followed in this study. The study also provides a comparative analysis of JKB with respect to other Burushaski varieties by studying its history and development since its split from them in 1890-1. It attempts to provide explanation of possible influences and their evaluation based on comparing the data collected in Srinagar with that of available literature on the Pakistani dialects as well as direct information gathered with the help of native speakers of these dialects.

Despite a high degree of multilingualism, the JKB community has successfully maintained its native language through several generations because of a strong sense of separate cultural identity and also a certain degree of social superiority among many owing to their royal descent. While it has preserved many relics of older forms lost in other dialects, as a result of 116 years of isolation from the major dialects and a unique language contact situation, JKB has developed various features which are systematically different from the other dialects of Burushaski. The changes and/or innovations have influenced all the subsystems of the language, viz. phonology, lexicon, morphology, and syntax. Some of the most significant of these changes are:

1. *Stress and intonation patterns:* JKB differs from other Burushaski dialects in terms of word stress and intonational patterns. Specifically, word stress in JKB is on the verge of becoming purely phonological, leaning towards a bias for word-initial primary stress based on Kashmiri as opposed to the inherent stress pattern which is morphologically-sensitive. Most of these cases involve syncopation of medial unstressed vowels just like Kashmiri. A pronounced similarity with Kashmiri is also observed with respect to intonational patterns.
2. *Loss of consonantal phonemes:* a gradual loss of retroflex sounds and their merger with the corresponding palatals is currently in progress in JKB. Thus, the retroflex fricatives *ṣ* and *ẓ*, the retroflex affricate *c̣*, and the retroflex glide *ɻ*, are being replaced by palatals *š*, *ž*, *č*, and *y* respectively.
3. *Re-establishment of lexical borrowings:* a large number of loanwords, initially phonologically nativized in Burushaski, are being re-borrowed through renewed contact with Kashmiri and Urdu. This has implications for some phonological changes in JKB (E.g., a change which involved word-final neutralization of voiced and voiceless consonants – all surfacing as voiceless consonants, no more applies on the borrowed vocabulary).
4. *Transfer of semantico-syntactic patterns:* a number of innovative linguistic constructions are found in JKB as a result of pattern transfer from Kashmiri and Urdu. Some of these changes are complex, involving simultaneous influence from both Kashmiri and Urdu. E.g., the case of pattern transfer of tag questions which involves both lexical borrowing from Urdu and pattern transfer from Kashmiri. The change has implications on the word order typology of JKB where the tag question forming function word *mat*, borrowed from Urdu, is intra-sentential occupying the second position in the sentence (as opposed to the inherent extra-sentential *be* or *na* – a common areal feature, except in Kashmiri. Kashmiri exhibits the so-called V-2 phenomenon where inflected verb occupies the second position in the sentence).
5. *Morphological simplification:* JKB inflectional morphology is being simplified in that the distinction between class features in terms of [+/-concrete] in [-human] nouns is in the process of being eliminated. Both nominal and verbal inflections are being affected by the change where the forms originally representing markers for [-concrete] are being lost.

(d) **Publications and Conference Presentations resulting from the NSF award:**

i. "Jammu and Kashmir Burushaski: Language, Language Contact and Change", University of Texas at Austin doctoral dissertation. (2006)
ii. "Contact-Induced Change in J & K Burushaski", paper presented at *Symposium About Language and Society*, Austin, University of Texas at Austin (April 2006)
iii. "Language-Contact and Change in J & K Burushaski", paper presented at the *South Asian Languages Analysis roundtable* – XXV, University of Illinois, Urbana-Champaign. (Sept. 2005)

(e) A brief description of the available data: The data gathered consist of elicited words and sentences and audio and/or video recordings of stories, personal narratives, poems, and naturally occurring conversations in Srinagar Burushaski (Refer to section 3.1 for more details).

(f) Relationship of the study to proposed project: The study is directly related to the proposed project as data from this study will significantly contribute to the projected outcomes.